

Aplicação de modelos de aprendizado de máquina para classificação de Fake News

Alexandre Augusto Foppa¹, Cristiane Machado¹, Marcelo Reis Bohrer¹

¹Universidade do Vale do Rio dos Sinos (UNISINOS)

alfoppa@hotmail.com, crism.stg@gmail.com, marceloreisbohrer@gmail.com

1. Introdução

No ambiente online as mídias sociais são um vasto campo para divulgação e propagação dos mais diversos conteúdos, e o que mais se destaca atualmente são as notícias falsas ou as chamadas *Fake News*. Além de ser o conteúdo mais acessado, as notícias falsas tem um grande potencial de circulação e podem ser postadas por qualquer pessoa ou entidade. Em razão disso, tais notícias devem, na medida do possível, ser verificadas quanto à sua autenticidade. Dentro deste contexto, o presente trabalho propõe mecanismos técnicos de classificação de notícias verdadeiras e falsas, ajudando a combater os danos causados pela disseminação da desinformação.

2. Definição dos objetivos

Baseados na problemática de *Fake News* e no conhecimento adquirido na disciplina de *Machine Learning*, foram pesquisados modelos utilizados para identificação de notícias falsas e reais. O objetivo desta modelagem é a formulação de um modelo de *machine learning* que auxilie na classificação de notícias.

O processo representado na Figura 1, foi especificado como forma de trabalho na fase de desenvolvimento do projeto

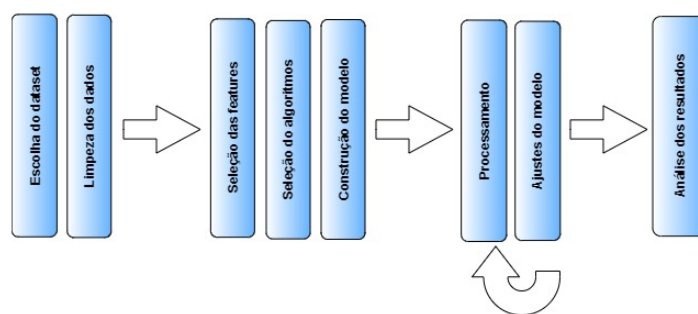


Figura 1. Metodologia de trabalho

3. Identificação e levantamento de dados

A escolha do dataset deu-se por meio de busca no site Kaggle, utilizando a palavra-chave “fake news”. Optamos por um dataset em língua inglesa, por conveniência, e, para o projeto, utilizamos os rótulos disponíveis no banco de dados e apenas o texto da mensagem, visando o desenvolvimento de uma aplicação que possa

ser integrada com um leitor HTML, independente do emissor ou canal de compartilhamento.

Utilizamos o dado disponibilizado na coluna “text_without_stopwords”, partindo do texto pré-processado sem palavras de ordem. A saída esperada do algoritmo é uma inferência booleana indicando se a notícia é verdadeira ou não, e/ou um escore de 0 a 1 indicando a probabilidade de ser verdadeira ou não.

4. Identificação de tipos de algoritmos para a abordagem

Dada a natureza textual do dado, será utilizado um algoritmo de Processamento de Linguagem Natural (PLN), utilizando um algoritmo para extração de features do texto da biblioteca *Scikit-learn*. Com as features extraídas, testaremos diferentes modelos para avaliar o que melhor atende a demanda, sendo eles *Decision Tree*, *Random Forest*, *Naive Bayes*, *Extra Trees* e *Logistic Regression*, também através da biblioteca citada. Além disso, testaremos uma *Artificial Neural Networks* (ANN), utilizando Keras e Tensorflow para avaliarmos o desempenho de um algoritmo de *deep learning*.

5. Conclusão

Com base na literatura, acreditamos ser possível as seguintes hipóteses:

- Com base nos dados disponíveis, o algoritmo terá uma eficácia aceitável para diferenciar notícias reais de falsas;
- Com base nos dados disponíveis, o algoritmo terá uma eficácia aceitável para identificar notícias reais, reduzindo a quantidade de notícias a ser analisadas, mas terá dificuldade de identificar notícias falsas [Edell 2018];
- Com base nos dados disponíveis, o algoritmo não terá uma eficácia aceitável para diferenciar notícias reais de falsas;

As duas primeiras hipóteses atendem o objetivo, sendo a primeira um cenário ideal, e a segunda um cenário mais realista de acordo com a literatura. Caso a terceira hipótese se comprovar verdadeira, serão buscadas alternativas, como utilização de outros tipos de dados disponíveis no dataset atual, ou buscar um dataset alternativo.

7. Referências

Edell, Aaron.(2018) “*I trained fake news detection AI with >95% accuracy, and almost went crazy*”, <https://towardsdatascience.com/i-trained-fake-news-detection-ai-with-95-accuracy-and-almost-went-crazy-d10589aa57c>, October.

Kaggle.(2020)https://www.kaggle.com/ruchi798/source-based-news-classification?select=news_articles.csv, October.

Shu, Kay, et. al.(2017) “*Fake News detection on Social Media: A Data mining perspective*”, <https://doi.org/10.1145/3137597.3137600>, New York, ACM, Volume 19, October

Scikit-learn.(2020) <https://scikit-learn.org/stable/modules/classes.html>, October.